Expanding the Range of Automatic Emotion Detection in Microblogging Text

Jasy Liew Suet Yan
School of Information Studies
Syracuse University
Syracuse, New York, USA
jliewsue@syr.edu

Abstract

Detecting emotions on microblogging sites such as Twitter is a subject of interest among researchers in behavioral studies investigating how people react to different events, topics, etc., as well as among users hoping to forge stronger and more meaningful connections with their audience through social media. However, existing automatic emotion detectors are limited to recognize only the basic emotions. I argue that the range of emotions that can be detected in microblogging text is richer than the basic emotions, and restricting automatic emotion detectors to identify only a small set of emotions limits their practicality in real world applications. Many complex emotions are ignored by current automatic emotion detectors because they are not programmed to seek out these "undefined" emotions. The first part of my investigation focuses on discovering the range of emotions people express on Twitter using manual content analysis, and the emotional cues associated with each emotion. I will then use the gold standard data developed from the first part of my investigation to inform the features to be extracted from text for machine learning, and identify the emotions that machine learning models are able to reliably detect from the range of emotions which humans can reliably detect in microblogging text.

1 Introduction

The popularity of microblogging sites such as Twitter provide us with a new source of data to study how people interact and communicate with their social networks or the public. Emotion is a subject of interest among researchers in behavioral studies investigating how people react to different events, topics, etc., as well as among users hoping to forge stronger and more meaningful connections with their audience through social media. There is growing interest among researchers to study how emotions on social media affect stock market trends (Bollen, Mao, & Zeng, 2011), relate to fluctuations in social and economic indicators (Bollen, Pepe, & Mao, 2011), serve as a measure for the population's level of happiness (Dodds & Danforth, 2010), and provide situational awareness for both the authorities and the public in the event of disasters (Vo & Collier, 2013).

In order to perform large-scale analysis of emotion phenomena and social behaviors on social media, there is a need to first identify the emotions that are expressed in text as the interactions on these platforms are dominantly text-based. With the surging amount of emotional content on social media platforms, it is an impossible task to detect the emotions that are expressed in each message using manual effort. Automatic emotion detectors have been developed to deal with this challenge. However, existing applications still rely on simple keyword spotting or lexicon-based methods due to the absence of sufficiently large emotion corpora for training and testing machine learning models

Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics, pages 38–44,
Gothenburg, Sweden, April 26-30 2014.

(Bollen, Pepe, et al., 2011; Dodds & Danforth, 2010).

Research in using machine learning techniques to process emotion-laden text is gaining traction among sentiment analysis researchers, but existing automatic emotion detectors are restricted to identify only a small set of emotions, thus limiting their practicality for capturing the richer range of emotions expressed on social media platforms. The current state-of-the-art of simply adopting the basic emotions described in the psychology literature as emotion categories in text, as favored by a majority of scholars, is too limiting. Ekman's six basic emotions (happiness, sadness, fear, anger, disgust, and surprise) (Ekman, 1971) are common emotion categories imposed on both humans and computers tasked to detect emotions in text (Alm, Roth, & Sproat, 2005; Aman & Szpakowicz, 2007; Liu, Lieberman, & Selker, 2003). It is important to note that most basic emotions such as the six from Ekman are derived from facial expressions that can be universally recognized by humans. Verbal expressions of emotion are different from non-verbal expressions of emotion. Emotions expressed in text are richer than the categories suggested by the basic emotions. Also, people from different cultures use various cues to express a myriad of emotions in text.

By using a restricted set of emotion categories, many emotions not included as part of the basic set are ignored or worse still, force-fitted into one of the available emotion categories. This introduces a greater level of fuzziness in the text examples associated with each emotion.

Example [1]: *"My prayers go to family of Amb. Stevens & others affected by this tragedy. We must not allow the enemy to take another. http://t.co/X8xTzeE4"*

Example [1] is an obvious case of "sympathy" as the writer is expressing his or her condolences to people affected by a tragedy. If "sympathy" is not in the pre-defined list of emotion categories that humans can choose from, human annotators may label this instance as "sadness", which is not entirely accurate. These inaccuracies will then be propagated into the automatic emotion detector.

While the basic emotions have been established as universal emotions (Ekman, 1999), their usefulness in emotion detection in text is still unclear. How useful are the six basic emotions in detecting consumers' emotional reactions towards a product or service from microblogs? What if a company wishes to detect disappointment? The focus on only the basic emotions has resulted in a dearth of effort to build emotion detectors that are able to recognize a wider range of emotions, especially the complex ones. Complex emotions are not merely combinations of the basic ones. For example, none of the combinations of Ekman's six basic emotions seem to represent "regret" or "empathy". Without human-annotated examples of complex emotions, automatic emotion detectors remain ignorant of these emotions simply because they are not programmed to seek out these "undefined" emotions.

There is a need to create automatic emotion detectors that can detect a richer range of emotions apart from the six basic emotions proposed by Ekman to deal with emotional content from social media platforms. A broader range of emotions will enable automatic emotion detectors to capture more fine-grained emotions that truly reflect actual human emotional experience. Limited research has been done so far to determine the full range of emotions which humans can reliably detect in text, as well as salient cues that can be used to identify distinct emotions in text. A crucial step to address this gap is to develop a gold standard corpus annotated with a richer set of emotions for machine learning models to learn from.

My research goal is to first discover the range of emotions humans can reliably detect in microblogging text, and investigate specific cues humans rely on to detect each emotion. Is there a universal set of cues humans rely on to detect a particular emotion or do these cues differ across

individuals? Using grounded theory, the first part of my investigation focuses on discovering the range of emotions from tweets collected from a popular microblogging site, Twitter, and the emotional cues associated with each emotion. Twitter offers a wealth of publicly available emotional content generated by a variety of users on numerous topics. The inherently social nature of interactions on Twitter also allows me to investigate social emotions apart from personal emotions. In the second part of my investigation, human annotations from the first part of my investigation will serve as gold standard data for machine learning experiments used to determine the emotions that automatic methods can reliably detect from the range of emotions that humans can reliably identify.

2 Background

Early research on automatic emotion detection in text is linked to subjectivity analysis (Wiebe, Wilson, Bruce, Bell, & Martin, 2004; Wiebe, Wilson, & Cardie, 2005). Emotion detection in text is essentially a form of sentiment classification task based on finer-grained emotion categories. Automatic emotion detection has been applied in the domain of emails (Liu et al., 2003), customer reviews (Rubin, Stanton, & Liddy, 2004), children's stories (Alm et al., 2005), blog posts (Aman & Szpakowicz, 2007), newspaper headlines (Strapparava & Mihalcea, 2008), suicide notes (Pestian et al., 2012), and chat logs (Brooks et al., 2013). Early development of automatic emotion detectors focused only on the detection of Ekman's six basic emotions: happiness, surprise, sadness, fear, disgust, and anger (Alm et al., 2005; Aman & Szpakowicz, 2007; Liu et al., 2003; Strapparava & Mihalcea, 2008). Plutchik's model is an expansion of Ekman's basic emotions through the addition of trust and anticipation in his eight basic emotions (Plutchik, 1962), while Izard's ten basic emotions also include guilt and shame (Izard, 1971).

Scholars have only recently started to expand the categories for automatic emotion classification as noted in the 14 emotions that are pertinent in the domain of suicide notes (Pestian et al., 2012), and 13 top categories that are used for emotion classification out of 40 emotions that emerged from the scientific collaboration chat logs (Brooks et al., 2013; Scott et al., 2012). However, existing gold standard corpora are limited by the emotion categories that are most often specific to a particular domain. Furthermore, it is difficult to pinpoint the exact words, symbols or phrases serving as salient emotion indicators because existing gold standard data are manually annotated at the sentence or message level.

Using Twitter, scholars have explored different strategies to automatically harness large volumes of data automatically for emotion classification. Pak & Paroubek (2010) applied a method similar to Read (2005) to extract tweets containing happy emoticons to represent positive sentiment, and sad emoticons to represent negative sentiment. First, this limits the emotion classifier to detect only happiness and sadness. Second, the lack of clear distinctions between the concepts of sentiment and emotion is problematic because tweeters may express a negative emotion towards an entity which they hold a positive sentiment on, and vice versa. For example, a tweeter expressing sympathy to another person who has experienced an unfortunate event is expressing a negative emotion but the tweet contains an overall positive sentiment. Third, such a data collection method assumes that the emotion expressed in the text is the same as the emotion the emoticon represents, and does not take into account of cases where the emotion expressed in the text may not be in-sync with the emotion represented by the emoticon (e.g., sarcastic remarks).

Mohammad (2012) and Wang, Chen, Thirunarayan, & Sheth (2012) applied a slightly improved method to create a large corpus of readily-annotated tweets for emotion classification. Twitter allows the use of hashtags (words that begin with the # sign) as topic indicators. These scholars experimented with extracting tweets that contain a predefined list of

emotion words appearing in the form of hashtags. Mohammad (2012) only extracted tweets with emotion hashtags corresponding to Ekman's six basic emotions (#anger, #disgust, #fear, #joy, #sadness, and #surprise) while Wang et al. (2012) expanded the predefined hashtag list to include emotion words associated with an emotion category, as well as the lexical variants of these emotion words. Although this method allows researchers to take advantage of the huge amount of data available on Twitter to train machine learning models, little is known about the specific emotional cues that are associated with these emotion categories. Also, this data collection method is biased towards tweeters who choose to express their emotions explicitly in tweets.

Kim, Bak, & Oh (2012) proposed a semi-supervised method using unannotated data for emotion classification. They first applied Latent Dirichlet Allocation (LDA) to discover topics from tweets, and then determined emotions from the discovered topics by calculating the pointwise mutual information (PMI) score for each emotion from a list of eight emotions given a topic. The evaluation of this method using a corpus of manually annotated tweets revealed that this automatic emotion detector only managed to correctly classify 30% of tweets from the test dataset. The gold standard corpus used for evaluation was developed through manual annotations using Amazon Mechanical Turk (AMT). Only 3% of the tweets received full agreement among five annotators.

3 Defining Emotions In Text

In everyday language, people refer to emotion as prototypes of common emotions such as happiness, sadness, and anger (Fehr & Russell, 1984). In the scientific realm, emotion is generally defined as "ongoing states of mind that are marked by mental, bodily or behavioral symptoms" (Parrott, 2001). Specifically, each emotion category (e.g., happiness, sadness, anger, etc.) is distinguishable by a set of mental, bodily or behavioral symptoms. When a person expresses emotion in text, these symptoms are encoded in written language (words, phrases and sentences).

Emotion in text is conceptualized as emotion expressed by the writer of the text. Emotion expression consists of "signs that people give in various emotional states", usually with the intention to be potentially perceived or understood by the others (Cowie, 2009). People express their emotional states through different non-verbal (e.g., facial expression, vocal intonation, and gestures) and verbal (e.g., text, spoken words) manifestations. Emotion expression in text is a writer's descriptions of his or her emotional experiences or feelings. It is important to note that emotion expression only provides a window into a person's emotional state depending on what he or she chooses to reveal to the others. It may not be depictions of a person's actual emotional state, which is a limitation to the study of emotion in text (Calvo & D'Mello, 2010).

4 Research Questions

Detecting emotions in microblog posts poses new challenges to existing automatic emotion detectors due to reasons described below:

- Unlike traditional texts, tweets consist of short texts expressed within the limit of 140 characters, thus the language used to express emotions differs from longer texts (e.g., blogs, news, and fairy tales).
- The language tweeters use is typically informal. Automatic emotion detectors must be able to deal with the presence of abbreviations, acronyms, orthographic elements, and misspellings.
- Emotional cues are not limited to only emotion words. Twitter features such as #hashtags (topics), @username, retweets, and other user profile metadata may serve as emotional cues.

Using data from Twitter, a popular microblogging platform, I will develop an initial framework to study the richness of emotions

expressed for personal, as well as for social purposes. My research investigation is guided by the research questions listed below:

- What emotions can humans reliably detect in microblogging text?
- What salient cues are associated with each emotion?
- How can good features for machine learning be identified from the salient cues humans associate with each emotion?
- What emotions in microblogging text can be reliably detected using current machine learning techniques?

5 Proposed Methodology

My research design consists of three phases: 1) small-scale inductive content analysis for code book development, 2) large-scale deductive content analysis for gold standard data development, and 3) the design of machine learning experiments for automatic emotion detection in text.

5.1 Data Collection

When sampling for tweets from Twitter, I will utilize three sampling strategies to ensure the variability of emotions being studied. First, I will collect a random sample of publicly-available tweets. This sampling strategy aims to create a sample that is representative of the population on Twitter but may not produce a collection of tweets with sufficient emotional content. The second sampling strategy is based on topics or events. To ensure that tweets are relevant to this investigation, tweets will be sampled based on hashtags of events likely to evoke text with emotional content. Topics will include politics, sports, products/services, festive celebrations, and disasters.

The third sampling strategy is based on users. This sampling strategy allows me to explore the range of emotions expressed by different individuals based on different stimuli, and not biased towards any specific events. To make the manual annotation feasible, I plan to first identify the usernames of 1) active tweeters with a large number of followers (e.g., tweets from politicians) to ensure sufficient data for analysis, and 2) random tweeters to represent "average" users of Twitter. I acknowledge that this sampling strategy may be limited to only certain groups of people, and may not be representative of all Twitter users but it offers a good start to exploring the range of emotions being expressed in individual streams of tweets.

5.2 Phase 1

To develop a coding scheme for emotion annotation, I will first randomly sample 1,000 tweets each from the random, topic-based, and user-based datasets for open coding. I will work with a small group of coders to identify the emotion categories from a subset of the 1,000 tweets. Coders will be given instructions to assign each tweet with only one emotion label (i.e., the best emotion tag to describe the overall emotion expressed by the writer in a tweet), highlight the specific cues associated with the emotion, as well as identify the valence and intensity of the emotion expressed in the tweet.

To verify the grouping of the emotion tags, coders will be asked to perform a card sorting exercise to group emotion tags that are semantically similar in the same group. Based on the discovered emotion categories, nuanced colorations within each category may be detected from the valence and intensity codes.

Coders will incrementally annotate more tweets (300 tweets per round) until a point of saturation is reached, where new emotion categories stop emerging from data. I will continuously meet with the coders to discuss disagreements until the expected inter-annotator agreement threshold for the final set of emotion categories is achieved.

5.3 Phase 2

Using the coding scheme developed from Phase 1, I will obtain a larger set of manual annotations using Amazon Mechanical Turk (AMT). AMT allows me to collect manual annotations of

emotions on a large-scale, thus enabling me to investigate if there are any differences as to what a larger crowd of people identify as emotion cues in tweets. Each tweet will be annotated by at least three coders. To ensure the quality of the manual annotations collected from AMT, workers on AMT will have to undergo a short training module explaining the coding scheme, and will have to pass a verification test before being presented with the actual tweets to be annotated. Inter-annotator agreement will be calculated, and the emotion categories that humans can reliably detect in text will be identified.

5.4 Phase 3

Detecting a single emotion label for each tweet can be defined as a multi-class classification problem. The corpus from Phase 2 will be used as training data, and the corpus from Phase 1 will be used as testing data for the machine learning model. An analysis of the emotional cues from Phase 1 and Phase 2 datasets is conducted to identify salient features to be used for machine learning. Support vector machines (SVM) have been shown to perform well in this problem space (Alm et al., 2005; Aman & Szpakowicz, 2007; Brooks et al., 2013; Cherry, Mohammad, & de Bruijn, 2012) so I will run experiments using SVM, and compare the performance of the model against a baseline using simple lexical features (i.e., n-grams).

6 Research Contributions

Analyzing the emotional contents in tweets can expand the theoretical understanding of the range of emotions humans express on social media platforms like Twitter. From a natural language processing standpoint, it is also crucial for the community to gain clearer insights on the cues associated with each fine-grained emotion. On top of that, findings from the machine learning experiments will inform the community as to whether training the machine learning models based on data collected using usernames, instead of topic hashtags will reduce noise in the data, and improve the performance of automatic emotion detection in microblogging texts.

The expected contributions of this research investigation are three-fold: 1) the construction of an emotion taxonomy and detailed annotation scheme that could provide a useful starting point for future research, 2) the creation of machine learning models that can detect a wider range of emotions in text in order to enable researchers to tap into this wealth of information provided by Twitter to study a greater multitude of behavioral and social phenomenon, and 3) findings on the range of emotions people express on Twitter can potentially help inform the design of social network platforms to be more emotion sensitive.

References

Alm, C. O., Roth, D., & Sproat, R. (2005). Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* (pp. 579–586). Stroudsburg, PA, USA.

Aman, S., & Szpakowicz, S. (2007). Identifying expressions of emotion in text. In *Text, Speech and Dialogue* (pp. 196–205).

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Bollen, J., Pepe, A., & Mao, H. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 450–453).

Brooks, M., Kuksenok, K., Torkildson, M. K., Perry, D., Robinson, J. J., Scott, T. J., … Aragon, C. R. (2013). Statistical affect detection in collaborative chat. Presented at the Conference on Computer Supported Cooperative Work and Social Computing, San Antonio, TX.

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, *1*(1), 18–37.

Cherry, C., Mohammad, S. M., & de Bruijn, B. (2012). Binary classifiers and latent sequence

models for emotion detection in suicide notes. *Biomedical Informatics Insights*, *5*, 147–154.

Cowie, R. (2009). Perceiving emotion: Towards a realistic understanding of the task. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1535), 3515–3525.

Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and Presidents. *Journal of Happiness Studies*, *11*(4), 441–456.

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation*, *19*, 207–283.

Ekman, P. (1999). Basic emotions. In *Handbook of Cognition and Emotion* (pp. 45–60). John Wiley & Sons, Ltd.

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, *113*(3), 464–486.

Izard, C. E. (1971). *The face of emotion* (Vol. xii). East Norwalk, CT, US: Appleton-Century-Crofts.

Kim, S., Bak, J., & Oh, A. H. (2012). Do you feel what I feel? Social aspects of emotions in Twitter conversations. In *International AAAI Conference on Weblogs and Social Media (ICWSM)*.

Liu, H., Lieberman, H., & Selker, T. (2003). A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th International Conference on Intelligent User Interfaces* (pp. 125–132).

Mohammad, S. M. (2012). #Emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*. Montreal, QC.

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Seventh International Conference on Language Resources and Evaluation (LREC)*.

Parrott, W. G. (2001). *Emotions in social psychology: Essential readings* (Vol. xiv). New York, NY, US: Psychology Press.

Pestian, J. P., Matykiewicz, P., Linn-Gust, M., South, B., Uzuner, O., Wiebe, J., … Brew, C. (2012). Sentiment analysis of suicide notes: A shared task. *Biomedical Informatics Insights*, *5*(Suppl. 1), 3–16.

Plutchik, R. (1962). *The Emotions: Facts, theories, and a new model*. New York: Random House.

Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop* (pp. 43–48). Stroudsburg, PA, USA.

Rubin, V. L., Stanton, J. M., & Liddy, E. D. (2004). Discerning emotions in texts. In *The AAAI Symposium on Exploring Attitude and Affect in Text (AAAI-EAAT)*.

Scott, T. J., Kuksenok, K., Perry, D., Brooks, M., Anicello, O., & Aragon, C. (2012). Adapting grounded theory to construct a taxonomy of affect in collaborative online chat. In *Proceedings of the 30th ACM International Conference on Design of Communication* (pp. 197–204). New York, USA.

Strapparava, C., & Mihalcea, R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 ACM Symposium on Applied Computing* (pp. 1556–1560). New York, USA.

Vo, B.-K. H., & Collier, N. (2013). Twitter emotion analysis in earthquake situations. *International Journal of Computational Linguistics and Applications*, *4*(1), 159–173.

Wang, W., Chen, L., Thirunarayan, K., & Sheth, A. P. (2012). Harnessing Twitter "big data" for automatic emotion identification. In *2012 International Conference on Privacy, Security, Risk and Trust (PASSAT), and 2012 International Conference on Social Computing (SocialCom)* (pp. 587–592).

Wiebe, J. M., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. *Computational Linguistics*, *30*(3), 277–308.

Wiebe, J. M., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, *39*(2-3), 165–210.